DOCUMENT RESUME

ED 236 357 CE 037 405

TITLE	Draft of a Model for Vocational Student Assessment.
INSTITUTION	Southern Association of Colleges and Schools, Atlanta, Ga.
SPONS AGENCY	Georgia State Dept. of Education, Atlanta. Office of Vocational Education.
PUB DATE	30 Apr 83
NOTE	65p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Academic Achievement; Behavioral Objectives; Criterion Referenced Tests; Guidelines; *Models; *Performance Tests; Postsecondary Education; Questionnaires; Secondary Education; State Programs; Statewide Planning; *Student Evaluation; Task Analysis; *Test Construction; Test Format; Test Items; Test Reliability; Test Use; Test Validity; *Vocational Education; *Vocational Evaluation
IDENTIFIERS	Assessment Systems; *Georgia

ABSTRACT

This volume contains the initial draft of a model for assessing students in vocational education programs in Georgia. Addressed in the first section of the draft are some of the components that are believed to be critical in the development of a model for assessing vocational student achievement, including selecting a program for use in developing the model; verifying duties, tasks, performance objectives, and guides; processing tasks; adopting or adapting tests; developing criterion-referenced tests (CRTs); editing written or performance tests; and administering tests. The second section of the draft, consisting of some of the formats and information necessary for the use of the model, deals with topics such as criteria for selecting an occupational area, developing a questionnaire for use in verifying tasks, specific topics that are covered on given existing vocational achievement tests, standards for evaluating CRTs, procedures for defining testing specifications, and guidelines for developing testing administrative manuals. (MN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED236357

DRAFT OF A MODEL

for

VOCATIONAL STUDENT ASSESSMENT

Submitted for Review

to the

Georgia Research Coordinating Unit

Office of Vocational Education

by

SOUTHERN ASSOCIATION OF COLLEGES AND SCHOOLS

B. E. Childers, Executive Director

Ronald McCage, V-TECS Director

K. M. Eaddy, Project Coordinator

CE037405

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

April 30, 1983

Atlanta, Georgia

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A Monchson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CONTENTS

SECTION I

		Page
I.	INTRODUCTION	1
II.	MISSION	4
III.	SELECT A PROGRAM FOR USE IN DEVELOPING THE MODEL	4
IV.	VERIFY DUTIES, TASKS, PERFORMANCE OBJECTIVES AND GUIDES	5
V.	PROCESS TASKS	8
VI.	ADOPT OR ADAPT TEST	9
VII.	DEVELOP CRITERION-REFERENCED TESTS	9
VIII.	EDIT EACH WRITTEN OR PERFORMANCT TEST	12
IX.	ADMINISTRATION	13

SECTION II

INFORMATION FOR ELEMENT 1.1	16
EXAMPLE OF FORMAT FOR ELEMENT 2.6	18
EXAMPLE FOR ELEMENT 2.13.1	19
EXAMPLE FOR ELEMENTS 2.13.2, 3, AND 4	20
INFORMATION AND EXAMPLE FOR ELEMENT 3.1	22

Page

INFORMATION FOR ELEMENTS 3.2 - 3.4 22

EXAMPLE FOR ELEMENT 4.1 23

EXAMPLE FOR ELEMENT 4.2 27

EXAMPLE FOR ELEMENT 5.1 - 5.4 30

EXAMPLE FOR ELEMENT 5.5 31

EXAMPLE FOR SUB-ELEMENT 5.6.1 32

INFORMATION FOR SUB-ELEMENT 5.6.2 36

INFORMATION FOR SUB-ELEMENT 5.6.3 36

INFORMATION FOR SUB-ELEMENT 5.6.4 38

INFORMATION FOR SUB-ELEMENT 5.6.5 40

INFORMATION FOR SUB-ELEMENT 5.6.6 41

INFORMATION FOR SUB-ELEMENT 5.6.7 AND 5.6.8 41

INFORMATION FOR SUB-ELEMENT 5.6.9 43

INFORMATION ON SUB-ELEMENT 5.6.9 and 5.8.1 AND 2 43

INFORMATION FOR SUB-ELEMENT 5.6.10 46

INFORMATION FOR SUB-ELEMENT 5.6.11 47

Page

INFORMATION FOR SUB-ELEMENT 5.7.3 49

INFORMATION FOR SUB-ELEMENT 5.7.6 49

INFORMATION FOR SUB-ELEMENT 5.7.7 50

INFORMATION FOR SUB-ELEMENT 5.7.8 51

INFORMATION FOR SUB-ELEMENT 5.7.9 53

INFORMATION FOR ELEMENT 5.9 54

INFORMATION FOR SUB-ELEMENT 5.7.10 57

NOTES ON ELEMENT 7.4 58

SECTION 1

I. INTRODUCTION

Models can be designed "for" or "of" a specific activity. The assignment of the Southern Association of Colleges and Schools was one "for" a model vocational student assessment system. Achievement of the assignment has not been conducted in isolation. The products of other persons working in the areas of performance testing, criterion-referenced test construction, task analysis, objective testing, and development of theoretical constructs for competency-based education have been examined. Many insights into the systems reviewed will undoubtedly be reflected in the resulting model "for" student assessment. Of particular influence will be the efforts of the Florida Student Assessment program, the American Institutes of Research Competency Measures, the Georgia Domain Assessment study, the Georgia Task Linkage project, and the numerous curriculum developments that utilize outcomes of the Vocational-Technical Education Consortium of States and task linkage projects.

The purpose of this draft model is to provide an opportunity for persons in the Office of Vocational Education and Southern Association of Colleges

and Schools to examine the components, and elements of the components, for comprehensiveness and logic. Therefore, the first section of this draft will depict those components believed, by the writer at this point, to be critical in the development of a model for assessing vocational student achievement.

Section II of this draft contains some of the formats and information necessary for use of the model. At this point they are not proposed as final instruments or complete information. They are; however, to be polished and used during the next four months in a pilot test of the model.

In developing a model the general nature of assessment has to be planned and specific objectives in regard to the ability or achievement to be measured must be determined.

In Georgia it is evident that much of the general nature has been provided in the Georgia State University task linkage project and the broad curriculum efforts conducted at the University of Georgia. In those comprehensive studies the purposes of assessment and the relationships of assessments to their intended uses have received adequate treatment. For instance, it is clear that the utility of the assessment results in vocational education

has multiple purposes; including, student achievements, evaluation of instructional programs, student placement, certification of learners, guidance, and evidence of completion of graduation requirements. The uses are directly related to the components of competency-based vocational education in which the following characteristics are empnasized.

1. Knowledge and skills must have a clear relationship to business and industry requirements for successful job performance. They are the foundations for vocational courses or programs.

2. The knowledge and skills are made public to parents, students, employers, and other publics at large.

3. The criteria for measuring success over the skills and knowledge are available to the students at the time of entry into a course or program.

4. Individualization of instruction is required. The emphasis is placed on making certain that each student achieves the stated tasks, under the conditions, and to the criteria specified in statements of course or program objectives.

While the general nature may be established for assessment, it is not so evident that the objectives to be measured in student assessment have

been adequately specified or that the assessment model has been totally developed. It is felt to be necessary, therefore, to show the components and elements of an ideal model, in draft form.

II. MISSION

To provide by September 30, 1984 a model for assessing vocational student achievement and a complete set of computerized test items over the objectives contained in one selected vocational program or course based on V-TECS catalogues. Test items are to be produced as a first step in developing objective and performance tests using the criterion-referenced objective test mode.

III. SELECT A PROGRAM FOR USE IN DEVELOPING THE MODEL

Component 1.0 Select a program for use in developing the model.

Elements:
- 1.1 State criteria for use in program selection
- 1.2 Obtain list of high priority vocational assessment areas from Georgia Office of Vocational Education

1.3 Obtain list of high priority vocational assessment areas from V-TECS members

1.4 Compare lists to show matches and mismatches

1.5 Discard mismatches

1.6 Examine matches over criteria

1.7 List those meeting criteria

1.8 Make selection (Georgia RCU staff).

IV. VERIFY DUTIES, TASKS

Component 2.0 Verify duties, tasks, performance objectives and guides, tool and equipment lists.

Elements:

2.1 Compile list of duties, tasks, objectives, guides, tools, and equipment from among task listing projects (if available), V-TECS catalogs, and curriculum materials

2.2 Analyze for entry level

2.3 Analyze for measurability

2.4 Analyze for enabling objectives

2.5 Analyze for problems of measurement; i.e., too broadly stated, no performance noted, no conditions listed, or implied

2.6 Prepare verification survey forms

2.7 Obtain list of incumbents, teachers, and supervisors to be used in verification study from RCU

2.8 Write letters of invitation to persons identified; including description of work to be performed, and acceptance form to be returned to RCU

2.9 Distribute survey forms with complete directions for use to persons conducting the studies of verification

2.10 Notify verification participants of meeting.

2.11 Hold meetings

2.11.1 Welcome participants

2.11.2 Explain procedure to be followed in completing the forms

2.11.3 Distribute surveys

2.11.4 Assist, when requested, by answering questions on procedure

2.11.5 Collect surveys and return unused and completed surveys to SACS project coordinator

2.11.6 Thank and dismiss participants.

2.12 Write letters of appreciation to participants with cover letters to administrators or supervisors

2.13 Analyze results of the verification study

2.13.1 Assign numerical ratings to each column in survey form (see example)

2.13.2 Summarize all ratings from all participants (see example)

2.13.3 Compute averages for supervisors and encumbents (see example)

2.13.4 List tasks by weighted importance, frequency and criticality (see example)

2.14 Report results of analysis

2.14.1 Select all tasks having an overall rating of at least 5.0

2.14.2 List task from high to low on separate sheet for submission to Office of Vocational Education representatives.

V. PROCESS TASKS

Component 3.0 Process Tasks

Elements:

3.1 Review each task for presence of action, object, and qualifing phrase

3.2 Review each task for presence of performance guides and objectives

3.3 Review each task for presence of tool and equipment list necessary to perform the tasks

3.4 Review each task for placement into a duty or collection of tasks describing a group of closely related activities.

VI. ADOPT OR ADAPT TEST

Component 4.0 Review tests for adoption, adaption

Elements:

4.1 Collect tests from A.I.R., SOCAT, Florida Department of Education, and others

4.2 Compare test items with criteria for evaluating CRT's

4.3 List those tests meeting criteria (to be adopted)

4.4 List those tests having similar requirements to the desired test

4.5 Discard tests not meeting criteria and list those as tests to be developed.

VII. DEVELOP CRITERION-REFERENCED TESTS

Component 5.0 Develop Criterion-Referenced Tests

Elements:

5.1 Secure verified list of course duties, objectives, performance guides, tools and equipment

5.2 Determine the category of test items required for each objective (written, performance, or both)

5.3 If written test is required, identify if knowledge or information is to be "selected" or "supplied" by student

5.4 If a performance test is required, then identify if "levels" of performance, or "acceptable" or "unacceptable" student responses are appropriate

5.5 Develop test specifications

5.6 Develop written test items for objectives requiring them

5.6.1 Develop an item budget

5.6.2 Assign test items

5.6.3 Review and modify items

5.6.4 Balance the answer key

5.6.5 Prepare test administration instructions

5.6.6 Review and revise (2nd)

5.6.7 Conduct a pilot test

5.6.8 Revise on basis of pilot test

5.6.9 Conduct a field test

5.6.10 Analyze field test findings, including reliability and validity data

5.6.11 Revise the test on basis of findings

5.7 Develop performance test items for objectives requiring them

5.7.1 Review test specifications (see 5.5)

5.7.2 Identify purposes, or uses to be made of the tests

5.7.3 Write instructions of exactly what examiners are to do. Step by step in the performance testing

5.7.4 List required tools, equipment, and materials

5.7.5 Write explicit instructions for examinee

5.7.6 Design a document for rating each examinee over each competency being measured

5.7.7 Select the tasks to be tested

5.7.8 Structure the tasks

5.7.9 Write the items

5.7.10 Review the items with subject matter experts

5.7.11 Test the performance test

5.8 Conduct validity study of test

5.8.1 Determine item-objective congruence

5.8.2 Determine construct validity

5.9 Determine reliability of criterion-referenced tests

5.10 Revise tests as necessary.

VIII. EDIT TEST

Component 6.0 Edit each written and performance test

Elements:

6.1 Collect all tests

6.2 Arrange written and performance tests

6.3 Provide final edit for grammar, style, length, coding, and required answer sheets

6.4 Produce clean copies for placement into computer or manual files.

IX. ADMINISTRATION

Component 7.0 Test Administration

Elements:

7.1 Write clear directions for test administrators and testees

7.2 Collect evidence of usefulness during pilot and field tests

7.3 Make revisions as required

7.4 Establish a clearinghouse for criterion-referenced tests to handle reproduction, distribution, processing, and reporting of test results.

SECTION II

SECTION II

Notes, Information, and Examples on Elements and Sub-Elements

of

Section I

Information for Element 1.1

Criterion 1. An occupational area should be selected from the task linkage projects. This criterion is necessary because of the relationships established in the task linkage between secondary and postsecondary programs. Also, the tasks have been organized and listed in a logical sequence believed to be important by Georgia educators for appropriate articulation.

In keeping with the requirement to place items in a computerized item bank, the coding of tasks by D.O.T. and O.E. numbers will be added useful information from the task linkage projects.

Criterion 2. An occupational area should be selected that has high interest for Vocational-Technical Education Consortium of States members. This is a requirement of the project, as approved by the State Board of Education, and therefore, cannot be discarded.

Criterion 3. An occupational area should be selected in which state or national licensure is not a requirement for employment.

Criterion 4. An occupational area should be selected for a technical area to be in keeping with the emphasis placed on expanding technology.

Criterion 5. An occupational area should be selected for which criterion-referenced test items are not presently available from some other source. Since many test items have been written and may be adopted or adapted, it is important to make thorough searches to identify available sources for possible use.

Example of Format for Element 2.6

Draft Verification Survey Form

Secondary cluster? Yes___ No___ Postsecondary program? Yes___ No___

Occupation_____D.O.T._____ _____ _____O.E. Code_____

List of tasks	Frequency performed				Importance			Should be learned	
	not done	weekly	monthly	quarterly	high	moderate	low	before or after job entry	
	___	___	___	___	___	___	___	___	___

Occupational title _____. Length ot time employed in months _____

Directions: Answer the questions about level of occupation with a check ✓ in the appropriate space. Fill in the occupation title, D.O.T. and O.E. codes. Then, complete the survey form by placing a check ✓ in the appropriate space to the right of each task as it relates to frequency, importance, and time to be learned. It is not necessary to supply your name. However, it is important that you indicate your occupational title and show the length of your employment in months.

Example for Element 2.13.1

TASKS	FREQUENCY				IMPORTANCE			
General Tasks	Not Done	Daily or Weekly	Monthly	Seldom or Quarterly	High	Moder-ate	Low	
(1) Greet guests		✓(3)			✓(4)			=7
2 Arrange tours for guests				✓(1)			✓(0)	=1
3. Schedule shifts for desk clerks	✓(0)							=0
Processing Guests								
35 Verify guest reserva-tions		✓(3)			✓(4)			=7
36 Arrange payments from guests		✓(3)			✓(4)			=7
(37) Check out Guests		✓(3)				✓(2)		
(38) Follow up unpaid bills			✓(2)			✓(2)		=4

Source: VECS materials, East Central Network for Curriculum Coordination

Example for Elements 2.13.2, 3, and 4

TASKS

General Tasks	Salience, Supervisors (Average)	Salience, Employees (Average)	Critical, (at least one mention)	Overall Rating (Supervisors and Employees)
1 Greet guests	7.0	7.0	11	7.0
2 Arrange tours for guests	2.8	0	-	1.9
3 Schedule shifts for desk clerks	2.3	1.0	-	1.9
Processing Guests				
35 Verify guest reservations	6.0	6.6	8	6.2
36 Arrange payments from guests	7.0	7.0	12	7.0
37 Check out guests	6.3	6.4	11	6.3
38 Follow up unpaid bills	4.0	1.0	-	3.0

Averages must be computed for the total number of supervisors and for the total number of employees responding to each item. Greater weights were assigned

to the ratings of supervisors but tasks rated very high by employees and only moderately high by supervisors were still included. Therefore, the following formula to arrive at an overall rating for each of the tasks was used:

$$R = \frac{2S + E}{3}$$

R = overall rating
S = supervisors' average
E = employees' average

For example, supervisors gave an average reaponse of 6.0 to Item 35, Verify Guest Reservations; on the same 7-point scale, employees rated it 6.6. Using the formula,

$$R = \frac{2(6) + 6.6}{3} = \frac{12 + 6.6}{3} = \frac{18.6}{3} = 6.2.$$

When this is done for each item on the inventory questionnaire, the result is a list of tasks and their weighted importance.

Source: VEMS materials

Information and Example for Element 3.1

Total task statements describe a job. In writing the statements, shown as an ordered list of activities, it is important to state them in simple, easy to understand, clear sentences. Vagueness has no place in task statements. For example, "preparing a financial statement" may be too broad a statement and summarizing the credits column may be too detailed a statement.

The following example will assist the reader to know what is expected in element 3.1.

Activity	Object	Qualifying phrase
Prepares	an income statement	using the Federal guidelines in F.R. 2958; 1981.

Information for Elements 3.2 - 3.4

Reference V-TECS guidelines and directions. Often it is a matter of checking against the content of a V-TECS catalog.

Example for Element 4.1

Test to be reviewed for adoption or adaption

AREA/TEST	* AIR	N	FL
AGRICULTURE:			
Agricultural Chemicals Applications			
Technician	X		
Agricultural Mechanics		X	
Farm Equipment Mechanic	X		
Greenhouse Grower			X
Horticulture		X	
Nursery Worker			X
Tractor Mechanic			X

(continued)

* AIR - The American Institutes for Research Vocational Competency Measures Project

N - NOCTI Student Occupational Competency Achievement Testing

FL - Florida Occupational Achievement Tests

Source: VECS Materials, East Central Network for Curriculum Coordination, Sangamon State University, Springfield, Illinois, 1982.

Element 4.1 - continued

AREA/TEST	AIR	N	FL
BUSINESS AND OFFICE:			
Accounting/Bookkeeping		X	
Clerk Typist			X
Computer Operator	X		
General Office		X	
Secretary (includes Stenographer module)			X
Word Processing Specialist	X		
DISTRIBUTIVE EDUCATION:			
Apparel Sales	X		
Fabric Sales	X		
General Merchandising		X	
Grocery Clerk	X		
Hotel (Motel) Front Office	X		
Salesperson, Parts			X

(continued)

Element 4.1 - continued

AREA/TEST	AIR	N	FL
HEALTH:			
Dental Assistant	X		
Hospital Ward Clerk			X
Nursing Aide			X
Physical Therapist Assistant	X		
Practical Nursing Occupations		X	
HOME ECONOMICS:			
Child Care Worker			X
Commercial Cook			X
Custom Sewing	X		
Fashion Construction Services		X	
Restaurant Service (Waiter, Waitress, Cashier)	X		
TECHNICAL:			
Electronics Technician	X		
Wastewater Treatment Technician	X		
Water Treatment Technician	X		

(continued)

Element 4.1 - continued

AREA/TEST	AIR	N	FL
TRADE AND INDUSTRY:			
Auto Body Repair		X	
Auto Mechanics		X	
Bricklayer		X	
Bricklayer (Construction)			X
Carpenter	X		
Construction Electricity (House)		X	
Diesel Mechanic	X		
Drafting		X	
Heating and Air Conditioning		X	
Industricl Electronics		X	
Machine Trades		X	
Plumbing		X	
Printing		X	
Radio/Television Repair		X	
Refrigeration		X	
Small Engine Repair		X	
Welding		X	

Example for Element 4.2

21 Standards for Evaluating CRTs

IB. MEASUREMENT PROPERTIES: CONCEPTUAL VALIDITY

* 1. Description: How good (i.e., thorough and understandable) are the descriptions of the objectives to be tested?

* 2. Agreement: How well do the test items match course objectives?

3. Representativeness: How adequately do the items sample objectives?

IB. MEASUREMENT PROPERTIES: FIELD TEST VALIDITY

4. Sensitivity: Does the publisher claim that conventional instruction leads to test score gains?

5. Item Uniformity: How similar are the scores provided by the publisher, on the different items for an objective?

6. Divergent Validity: Are the scores for each objective relatively uninfluenced by other skills?

7. Lack of Bias: Are test scores unfairly affected by social group factors?

8. Consistency of Scores: Are scores on individual objectives consistent over time or over parallel test forms?

Example for Element 4.2 - continued

II. APPROPRIATENESS FOR EXAMINEES

9. Clarity of Instructions: How clear and complete are the instructions to students?

10. Item Review: Does the publisher report that items were reviewed or field tested for quality?

11. Visible Characteristics: Is the layout and print easily readable?

12. Ease of Responding: Is recording answers a problem for pupils?

III. PRACTICALITY

13. Informativeness: Does the test buyer have adequate information about the test before buying it?

14. Curriculum Cross-Referencing: Are the test objectives indexed to at least two series of relevant teaching materials?

15. Flexibility: Are many of the objectives tested at more than one level, and are single objectives easy to test separately?

16. Alternate Forms: Are parallel forms available for each test?

* 17. Test Administration: Are the directions to the tester clear, complete, and easy to use?

Example for Element 4.2 - continued

18. Scoring: Are both machine scoring and easy hand scoring available?

19. Record Keeping: Does the publisher provide record forms keyed to test objectives that are easy to use?

20. Decision Rules: Are well justified, easy to use rules given for making instructional decisions on the basis of test results?

21. Comparative Data: Are scores of a representative reference group of students given for comparing with your students' scores?

* Critical factors to be considered by reviewers.

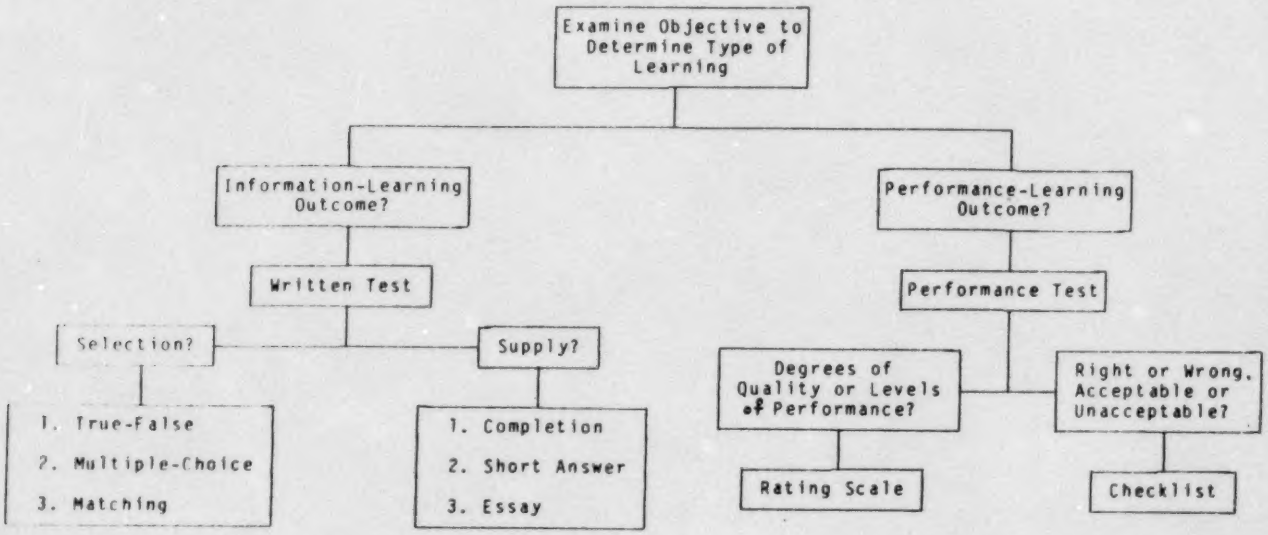
Source: Center for the Study of Evaluation, UCLA Graduate School of Education, Los Angeles, California

Example for Element 5.2 - 5.4

When you are ready to develop a test, examine each objective carefully to determine:

1. Does the objective require a student to determine knowledge or information about something?
2. Does the objective require the student to demonstrate the actual performance of a skill:

If Number 1 is answered in the affirmative then a written test is required. If Number 2 question is affirmative, then a performance test is required. The diagram below will help the reader.



Source: "Criterion Referenced Testing," Florida State Department of Education, 1981, p. 50.

Example for Element 5.5

The initial test specifications serve as the general blueprint. In designing this blueprint the test developer must consider the purpose of the test and the limitations of the "environment" in which the test will be used. Some limits that should be considered are:

- The amount of time that can reasonably be expected for administering a test to examinees
- The availability of equipment and materials for performance testing
- The grade or mastery level of the typical examinee

The topics covered in an outline of test specifications are:

- The types of measures to be used (paper-and-pencil, performance, or both) and their formats
- The total number of items in the finished form
- Total testing time
- The skill level to be assessed by the test
- General reading level of the instructions and questions

As the test is developed, various changes are likely to be made, but the specifications should provide guidance and coherence to a test development project.

Source: VECS materials printed by East Central Network for Curriculum Coordination, Sangamon State University, Springfield, Illinois, 1982.

Example for Sub-Element 5.6.1

The "budgeting of items" is a procedure for determining the number of paper-and-pencil test items to develop within each major test area. Your decisions about an item budget must be based on your best estimate; no hard and fast rules exist.

The first step is to establish the content of the item budget, that is, what will be the areas of an occupation to be covered in the test. The content is derived from the task inventory findings listing the skills and abilities that employers and employees consider important.

The next step is to estimate the number of items for the final test.

For example, if between 50 and 60 test items are wanted for the final version, at least double that number of items should be prepared initially. The test developer must then assign a percentage of the total number of test items being developed to each duty. Examples of major categories derived from an auto mechanics task inventory could be "safety," "trouble-shooting," and "tune-up procedures."

If you decide to prepare 100 items initially, and the test will have five duties, you could assign the same number to each category. In this case, you simply have:

$$100 \div 5 = 20$$

Often, some duties may be more important to the job than others; you may wish to assign a larger proportion of items to those duties. The distribution, then, could resemble the following:

Duty	No. of Items Per Category
A	30
B	10
C	20
D	25
E	15
TOTAL	100

Once the numbers of test items to be developed are assigned, the number of items to be developed for each task within each major duty in the task list must be determined. If, for example, a major duty is allotted 30 questions, such as duty "A" above, and if the duty has five specific tasks, then the 30 questions must be assigned to these five tasks.

Often, tasks to be used as the basis for item development are rated by those surveyed as to their importance. If, for example, tasks were rated either "moderately important" or "very important," such that:

Assigned Weight		
1	=	moderately important
2	=	very important

then the first category could be assigned "1" and the second, "2." Assignment of test items per task using the following simple procedure could be followed. Total weight of all the tasks are summed.

Task	Importance		Assigned Weight
1	moderate	=	1
2	very	=	2
3	very	=	2
4	moderate	=	1
5	moderate	=	1
			7

Source: VECS materials printed by East Central Network for Curriculum Coordination, Sangamon State University, Springfield, Illinois, 1982.

Then, divide the allotted number of questions, 30, by the total weight, 7:

$$30 \div 7 = 4.3$$

The result, 4.3, is then multiplied by the weight of the task, which in this case is either 1 or 2.

Task	Assigned Weight		Weight Multiplier		No. of Questions Per Task
1	1	x	4.3	=	4.3
2	2	x	4.3	=	8.6
3	2	x	4.3	=	8.6
4	1	x	4.3	=	4.3
5	1	x	4.3	=	4.3

Since these results are not whole numbers, it is necessary to round to the nearest whole number to get the appropriate number of questions to develop per task, so that the final result is:

Task	No. of Questions Per Task
1	4
2	9
3	9
4	4
5	4
TOTAL	30

Source: VECS materials printed by East Central Network for Curriculum Coordination, Sangamon State University, Springfield, Illinois, 1982.

Information for Sub-Element 5.6.2

Based on the test item budget, assign each item writer a specific number of test items to write in specific categories. Supply each writer with detailed instructions regarding the nature of the task, specific requirements, and the format of the questions to be developed. These instructions should clearly reflect the design goals of the test developer.

Of course, items should be assigned to writers who are knowledgeable about the specific tasks. But it is a good idea to ask a second writer to submit items on the same tasks to vary the emphasis and point of view.

Information for Sub-Element 5.6.3

The test development staff should review each completed item for clarity, completeness, correct grammar, etc. Items that are unclear, unimportant, or have content problems requiring technical expertise should be discussed with the item writer. Occasionally, when different individuals prepare items for the same occupational task, duplicate or near duplicate items will be written. The best item can be kept and the other placed in reserve, or possibly even

combined with the two to produce a better test item.

Then compare the selected items to the item budget. It may be found that more items than are needed are in one area and too few in another. In areas of shortage, the test developer should work closely with the item writers to stimulate ideas for questions. A "brainstorming" session can be very productive. Once there is a close match with the budget, the items should be prepared in final form. The language and structure of the questions and their options should be logically and grammatically consistent. All the items should then be looked at as a unit so that extraneous cues can be removed. A common error is that one test item will cue the correct response in another item.

The items should then be reviewed by subject matter experts not involved in the actual writing. This may include vocational instructors or persons in industry not previously involved. Items that are factually in error or items in which a disagreement exists about the correct response must be removed. A good strategy is to have the experts "take the test." Although experts make errors, it should be possible to catch most of the distractors that may be multiple correct responses to the test item.

Source: VECS materials printed by East Central Network for Curriculum Coordination, Sangamon State University, Springfield, Illinois, 1982.

Information for Sub-Element 5.6.4

Balancing the key is one of the final stages in the preparation of multiple-choice test items. It ensures that each possible option has approximately an equal proportion of correct answers assigned. It helps to eliminate any kind of bias that test designers or item writers may have in the placement of correct answers. For instance, some item writers have a tendency to plan the correct response as one of the middle options.

The balancing procedure to use is very simple:

1. List separately the correct response for all five-choice items, four-choice items, and three-choice items.
2. Sum the number of five-choice, four-choice, and three-choice items separately.
3. Divide each sum by the number of choices for that category.
4. The result show how many A, B, C, etc., correct responses should be in each category.
5. Compare the results against the actual number of items in each category.
6. Rearrange the options, whenever feasible, to match the balanced key.

The following is an example of the balancing procedure, using five-choice items. The same procedrue is used for four and three-choice items.

Item Choice	Present Key	Adjustments Needed	Revised Key
A	6	+2	8
B	12	-3	9
C	18	-9	9
D	1	+7	8
E	5	+3	8
TOTAL	42	TOTAL	42

Number of items - number of choices per item = $42 \div 5 = 8.4$

Approximately 8 questions having each item choice. In this case since the total is 42, two item choices must have one extra each.

When balancing the key, options cannot be moved around at will. The option item ordering guidelines described above should be maintained.

Source: VECS materials printed by East Central Newwork for Curriculum Coordination, Sangamon State University, Springfield, Illinois, 1982.

Information for Sub-Element 5.6.5

The best test can be a poor assessment tool if the instructions for its administration are not clear and specific. One method of providing these instructions is to prepare an examiner's manual. This manual should contain all the information required for administering the test. The manual is also useful in supplying background information to aid the examiner in understanding the overall structure and purpose of the test. Listed below is the areas that might be covered in an examiner's manual.

- Purpose of the test
- Expected background of examiner
- Ways test results may be used
- Overview of the test
- Administration of the test
  - test format
  - time required
  - answer sheets
  - instructions for administering the test
  - suggestions for testing individuals with special needs

Information for Sub-Element 5.6.6

When the test is completed, the entire package should be reviewed by the members of the test development process. To assure useful reviews, detailed instructions should be supplied to each reviewer.

Test revision based on the comments of reviewers should be carefully performed to avoid introducing errors into the test. Any changes of factual content that seem questionable should be submitted to the item writers. Carefully review any area where there are differences of opinion. If it is not possible to agree on the correct answer, give serious consideration to deleting the item.

Source: VECS materials printed by East Central Network for Curriculum Coordination, Sangamon State University, Springfield, Illinois, 1982.

Information for Sub-Element 5.6.7 and 5.6.8

Once the test has been revised, it is ready for pilot testing. Pilot testing is a procedure used primarily to test the structure of the test, e.g., clarity of instructions, time estimates, etc., rather than the content. The procedure for conducting a pilot test is to have two or three vocational

training programs each administer the test to two or three students. The tests should each be at a different school with different examiners and students. A packet of the complete test instructions should be sent to each examiner, one or two weeks in advance.

The examiner should be informed that the test designer will be at the test site as an observer and to debrief the examiner and examinees. Make it very clear that the test is being tested and that any problems with the test are not a refelction on the examiner or student but rather on the test. Ask the examiner and the examinees to be frank.

During testing, the test designer will act as an observer and should not supply assistance or answer questions. It is appropriate, however, to stop and ask questions for clarification. Careful notes should be kept, noting the strong and weak points of test administration.

After completion of the test, the observer should review the test with the examiner and students. An item-by-item review is most desirable. On completion of the pilot tests, corrections should be made to the test based on the findings. The findings of pilot testing will help improve instructions and "packaging" of the test.

Information for Sub-Element 5.6.9

Field testing is used to determine the quality of the test items; including validity and reliability. The field test involves a large number of students in a number of schools. Field testing requires considerable pre-planning in that agreements must be obtained from the schools and instructors. There are no hard and fast rules regarding the size of the sample. Approximately 100 students spread over five schools is a reasonable target.

Source: VECS materials printed by East Central Network for Curriculum Coordination, Sangamon State University, Springfield, Illinois, 1982.

Information on Sub-Element 5.6.9 and 5.8.1 and 2

Probably the most important function of a field test is the determination of test validity, content and construct. It is necessary to establish content validity to address the content relevance of a test (i.e., the match between item content and the objectives the items are written to measure). Construct validations are required to assess the validity of test scores for accomplishing their intended uses.

Content validity involves the judgments of test items by content specialists. The question addressed is this: Is the format and content of an item appropriate to measure some part of the task or objective specification? A procedure for assessing item-objective match involves the use of a rating scale. The information needed is the mean and median ratings assigned to the items by a group of content specialists. An examination of the range of the ratings given each item provides an indication of the extent of agreement among the content specialists. It is also possible to determine the "closeness" of each judge's ratings to the median responses of the group of judges. When one or more of the judges is far out of line and there is evidence of carelessness on the part of the judge, for example, it may be best to eliminate the response and recalculate the statistics. A summary and analysis of the hypothetical ratings of nine judges on fourteen test items measuring four objectives is depicted below:

Objective	Test Item	Judge's Ratings 1	2	3	4	5	6	7	8	9	Mean	Median	Range
1	2	4	3	5	5	4	5	5	5	4	4.4	5	3
	7	4	2	5	5	5	5	5	4	5	4.4	5	4
	14	4	5	5	5	4	5	5	5	5	4.8	5	2
2	1	3	5	3	2	1	4	5	2	4	3.2	3	5
	3	3	1	4	4	3	4	4	3	3	3.2	3	4
	8	1	3	1	2	1	1	1	1	1	1.3	1	3
	13	1	3	2	1	1	2	1	2	4	1.8	2	3
3	2	4	5	5	4	5	5	5	5	5	4.8	5	2
	6	4	2	4	4	4	4	4	4	4	3.8	4	3
	12	5	3	5	5	5	5	5	5	5	4.8	5	3
4	5	4	3	5	5	4	5	5	4	5	4.4	5	3
	9	2	2	4	1	4	2	4	4	4	3.0	4	4
	10	1	3	1	2	1	1	1	1	1	1.3	1	3
	11	4	3	4	4	5	5	5	5	5	4.6	5	3
Judges discrepancies from median responses		9	24	2	10	6	4	4	3	3			

Source: R. A. Berk (Ed) Criterion-Referenced Measurement, The John Hopkins University Press; Baltimore, Md. 1982, p. 91.

Construct validity is supposedly possible with CRT's scores. The studies should begin with a definite statement of the proposed use of the test scores. A clearly stated use will provide direction for the kind of evidence that is worth collecting.

One method that seems worthy of investigation is the experimental method. Students are randomly assigned to one of two groups. The first group receives instruction; the second group does not receive instruction. Instruction is on content contained in an occupational task listing. If the treatment is effective, higher test scores by the experimental group would support the construct validity hypothesis.

Information for Sub-Element 5.6.10

Field testing will provide the test developer with a large amount of data to analyze. If possible, the developer should work closely with someone skilled in statistical analysis. Access to a computer is necessary at this time. One of the numerous statistical software packages in place will be used. Most are simple to use and can accommodate the level of statistical analysis required for item analysis purposes.

Information for Sub-Element 5.6.11

In order to decide which items to keep and which to discard, it is necessary to get statistical data on each item to answer the following questions:

1. How well does examinee performance on each item correlate with overall examinee performance on all items in a particular occupational area? How well does an item discriminate between those who perform well and those who perform poorly, overall?
2. What proportion of examinees answered an item correctly?

The prime concern is to select items that were performed well in the field test and to end up with a test that has about the same percentage of items in each section as specified in the item budget. The criteria for item selection should be threefold:

1. Content--The items should be representative of the performance area.
2. Difficulty--What was the proportion of examinees who missed an item? In competency tests, items should be included over the range of difficulty reflected in the task or objectives; not on an arbitrarily selected level of difficulty.

3. Discrimination--Do those who do well on the test generally answer the item correctly and those who do poorly overall generally answer the item incorrectly?

In analyzing the paper-and-pencil items, consider these points. Distractors that were chosen by most examinees might need to be modified or deleted. The correct item selected by all examinees should not be eliminated or changed. Correct items which are seldom answered "correctly" should be checked.

All items should be carefully reviewed by subject matter experts to determine whether there is a possibility of misinterpretation.

After [redacted] test items have been selected, make sure [redacted] the same item distribution as was used on the field test is available. If some areas do not have enough items, look again at those items not selected.

Generally, poor items should be removed rather than rewritten, but if the change is small the test developer may want to make it in order to have a useful item. If a major change is required, then the rewritten item should be considered a new item that has not been field tested.

Information for Sub-Element 5.7.3

To ensure that a performance test is administered the same way to all examinees, detailed administration instructions must be prepared. An examiner's manual should cover all the components of a performance test. It should also include suggestions for testing individuals with special needs. Though the instructions to the examiner should be detailed enough to prevent variation between tests, they should not be unnecessarily wordy.

Information for Sub-Element 5.7.6

Rating sheets on which the examiner will rate examinee performance should be easy to follow and at the same time permit a thorough evaluation of the examinee's performance. The rating form should allow an examiner to observe and rate the examinee on each checkpoint without missing any of the examinee's ongoing performance on the test.

On completion of the final revision, you should have a useful paper-and-pencil test meeting the design requirements established at the start of the project.

Source: VECS materials printed by East Central Network for Curriculum Coordination, Sangamon State University, Springfield, Illinois, 1982.

Information for Sub-Element 5.7.7

Performance tests are more difficult to construct and more expensive to administer than paper-and pencil tests. Therefore, performance tests should be developed to cover only important areas that cannot be tested adequately using a paper-and-pencil test. It is important to discuss each proposed area with subject matter experts. Some questions to ask in selecting tasks for a performance test are:

1. Can the skills required to perform the task only be assessed adequately with a performance test and not with a paper-and-pencil test?
2. Was the area rated as important by employers and employees in the task inventory survey?

3. Are the equipment or required materials available or easily obtainable at a training site?

4. Will the cost in consumable items per examinee be reasonable?

Information for Sub-Element 5.7.8

The list of components of a performance test indicate that the final package is much more complex than a paper-and-pencil test package. It is the designer's job to make all the parts fit together so that those using the test feel it is simple to set up, administer, and score.

First, the task should be divided into all its observable behaviors and products. For example, if an examinee is asked to replace the head gasket of an automobile engine, the task should be broken down and each step recorded. This includes even those actions that seem trivial. In the case mentioned, we would start with:

1. Examinee opens hood of car.

2. Determines model of engine.

3. Checks to see whether proper gasket is available.

The listing would be finished when the examinee closes the hood of the car. The list should be ordered as closely as possible to the way the task is typically performed. The test developers should then look for possible products resulting from the process. Such points as "heads torqued to proper specification," "gasket straight and not leaking" are potential products that can be measured.

From the list, those points that are (a) important to the job, (b) likely to be performed wrong, and (c) relatively independent of other actions required for performance of the task should be extracted and used as assessment checkpoints. The checkpoints should then be written as descriptions of the action, and in enough detail to ensure that an examiner can make a judgment about correctness of a behavior/action. Too much detail can make it difficult for an examiner to rate an action in an on-going work situation.

Information for Sub-Element 5.7.9

The procedures for pilot testing and field testing a performance test are basically the same as those used for a paper-and-pencil test.

The individual performance tasks should correlate positively--but not necessarily highly--with the job knowledge test. If one or more correlate very highly, that raises a question of the cost effectiveness of these performance measures. Another concern is if a performance test has a negative correlation with the paper-and-pencil test. The correlations among performance measures should also be reviewed. Ideally, they will have low positive correlations.

Within each performance test, checkpoints that are performed correctly by all or nearly all of the examinees should be carefully reviewed. The importance of the item to the occupation and the task should be considered. Checkpoints that are missed by all or nearly all examinees should be restudied to determine what is causing the problem.

As part of final revision, the test developer should spend time working on the appearance and layout of the test. Every effort should be made to reduce the amount of paper and paperwork. The final version of the test

should be clear, well-organized, and easily administered. At this point, the developer should have a complete test package that will give objective information to directors of vocational education, instructors, and students about the capability, strengths, and weaknesses of students and programs.

Information for Element 5.9

Reliability is concerned with the consistency of test scores. There are three different concepts of CRT's reliability.

1. Reliability of mastery classification decisions,
2. Reliability of criterion-referenced test scores, and
3. Reliability of domain score estimates.

This project will be concerned with reliability of mastery and non-mastery decisions over repeated testing of the same group.

For example, those data below show total scores of 10 students on two ten item tests.

	Total Score	
	Test 1	Test 2
Student		
1	9	7
2	8	8
3	7	7
4	7	4
5	7	3
6	6	7
7	6	7
8	6	5
9	6	4
10	5	6

N = 10

Minimum master score = 8

	Form 2		
Form 1	Mastery	Nonmastery	Total
Mastery	1	1	2
Nonmastery	0	8	8
Total	1	9	10

The above summary indicates that reliability (or consistency) can be shown as an index and easily determined by teachers without a great deal of trouble. For example, if 8 is the criterion score for mastery, student #2 is the only student classified as master/master (using tests 1 and 2). Student #1 is classified as master/nonmaster and therefore a 1 is placed in that cell. The remaining eight students are placed in the nonmaster/nonmaster cell. It can be observed easily that 9 out of 10 students are consistently classified over the two tests. Numerically the formula could be shown as:

$$\hat{p}o = 1/10 + 8/10 = 9/10 = 0.90$$

$\hat{P}o$ = proportion. 1/10 = mastery/mastery . 0/10 = nonmastery/nonmastery.

With large numbers of students, the test results can be input on magnetic tape and processed for ease in handling and reporting with a computer.

Information for Sub-Element 5.7.10

As with paper-and-pencil tests, performance tests should be reviewed by subject matter experts in industry and by vocational educators. The points that reviewers should keep in mind are listed below. If a paper-and-pencil test is developed, it would be most efficient to have both the paper-and-pencil and performance tests evaluated by the individuals at the same time. The reviewers can then look at the complete test package.

- Are all required supplies and materials listed? If not, what is missing?
- Are the instructions adequate?
- Are all the relevant topics for evaluation included? If not, what topics are missing?
- Are the topics for evaluation listed in the order in which they would be carried out by an examinee taking the test?
- Are the materials provided sufficiently clear and complete? If not, what should be added to make them more satisfactory?
- Are the approximate time limits indicated in the test satisfactory?
- How can the problem be improved?

Notes on Element 7.4

The establishment of a clearinghouse to be responsible for test production, packaging, distribution, monitoring, processing, and reporting of results is an extremely important consideration for the RCU personnel.

The undertaking is an expensive one if it is to be productive. Factors to consider include the availability of professional personnel in the area of C-R-M; the presence of computer capability; the background of success in the business of student assessment; and the long range plans and commitment of an institution to provide the maintenance, research, and development capabilities necessary for proper administration.

It seems, at this point, that the administration must be institutionally based to provide the human and material resources required for success. Georgia is fortunate in having more than one institution available.

Probably the question needing immediate attention is; how does a state support continuing development and administration of C.R. tests? There are several obvious answers; (a) leave assessment to the individual teachers and schools, (b) develop, with state and Federal funds, several tests each year for broad distribution, (c) initiate the establishment of a production

unit within the State Department of Education, (d) continue to use tests produced in other states and hope the Georgia program objectives match those tested for in the instruments purchased, or (e) become a partner with other states having a need to use CRT's in the production of testing devices; thereby, reducing costs and increasing available materials while at the same time influencing quality control by participating in the decision making processes of the collaborative effort. More options could be identified, each having some strengths and weaknesses. However, the greatest percent return on dollar investment seems to be the cooperative mode identified in number "e" above.

If the Office of Vocational Education is to have maximum impact upon instruction through student assessment, answers to the questions of assessment administration require a high priority among the policy makers so that organization, funding, personnel assignment, and facility assignments can be addressed. If policy decisions are made in those areas; then personnel responsible for developing and implementing procedures can provide a program of student assessment realistic to the needs of Georgia students, educators, employers, and other taxpayers. Now is the time to address the issues.